

Article

Multi-Scale Progressive Mineral Prospecting Prediction of Hidden Manganese Ore Deposits Based on Geological Big Data in Northeastern Guizhou, China

Kai Xu ^{1,2,3,4} , Jinliang Ma ¹, Chonglong Wu ^{1,2,3,4} and Chunfang Kong ^{1,2,3,4,*} 

¹ School of Computer Science, China University of Geosciences, Wuhan 430074, China; xukai@cug.edu.cn (K.X.); 1202511191@cug.edu.cn (J.M.); wcl@cug.edu.cn (C.W.)

² Engineering Technology Innovation Center of Mineral Resource Exploration in Bedrock Zones, Ministry of Natural Resources, Guiyang 550081, China

³ Guizhou Provincial Laboratory of Intelligent Prospecting and Exploration, Guiyang 550081, China

⁴ Hubei Key Laboratory of Intelligent Geo-Information Processing, Wuhan 430078, China

* Correspondence: kongcf@cug.edu.cn; Tel.: +86-18602766895

Abstract

Exploration for hidden ore bodies is important and requires new prospecting methods. However, predicting the presence of deep hidden mineral resources is difficult because of the large prospective area and the numerous factors influencing underground geological structures and mineral resources; in addition, available information on geological structures and their relationships and changes over time is incomplete. This study focused on a region with typical Mn ore deposits in the Upper Yangtze Block, northeastern Guizhou Province, South China. A geological model based on basic geology and ore mineralogy knowledge was combined with model-free prediction based on the fourth paradigm of scientific research—which uses the multi-scale progressive technical strategy from 1P to 5P (xP refers to various scales of mineralization potential prediction)—to generate a data-and-model-driven big data mineral prospectivity model (BDMPM) that organizes big datasets and big data chains to predict Mn ore mineralization potential. The validation of BOA-AdaBoost in the Songtao Mn ore-concentrated areas demonstrated that the prediction model at the 3P scale achieved an accuracy of 0.936, precision of 0.948, recall of 0.925, F1 score of 0.928, Kappa of 0.862, and AUC of 0.967. This also indicated that the proposed BDMPM can provide technical support for the prospecting and exploration of hidden Mn ore deposits.

Keywords: Datangpo-type manganese ore (DMO); big data mineral prospectivity model (BDMPM); multi-scale progressive prediction; data mining; correlation relationships



Academic Editor: Paul Alexandre

Received: 14 August 2026

Revised: 23 September 2026

Accepted: 24 September 2026

Published: 28 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Mineral resource prediction involves the systematic prediction and integrated evaluation of probable occurrences and endowment characteristics of economically exploitable mineral resources on Earth. In recent decades, with the rapid advancement of artificial intelligence (AI) technology, mineral prospectivity models (MPMs) have also undergone significant development [1–3]. Generally, MPMs can be classified into three main categories: those based on geological formations and mineral indicators (knowledge-driven), those derived from geological features extracted from data (data-driven), and those employing hybrid approaches [4,5].

Knowledge-driven MPMs are primarily based on statistical analysis techniques, such as fuzzy logic [6], index overlay [6], analytic hierarchy process [7], and multi-criteria

decision-making (MCDM) [8]. Knowledge-driven MPMs mainly depend on the empirical knowledge and experience of geologists, and their parameters are based on expert knowledge of the processes that lead to the formation of ore deposits in a given geological environment. This experiential knowledge enables them to identify and interpret the formation of new ore deposits. Knowledge-driven MPMs are typically used in areas with only a few or no known ore deposits, or poorly explored greenfields where exploration data are scarce [9,10].

By comparison, data-driven MPMs are based on using artificial intelligence (AI) technology to conduct quantitative analysis of the spatial correlation between ore-forming prediction indices and known ore deposits, and then using the established mathematical model to predict metallogenic potential [1,11–13]. In the development of AI systems, machine learning (ML) and deep learning (DL) techniques have made remarkable progress in data-driven MPM tasks [1,11]. These ML models include the weights-of-evidence [14–16], logistic regression [17,18], support vector machine [19], ensemble learning [10,20,21], random forest [5,10,13,21–23], decision tree analysis [10], and XGBoost [10,21]. The superiority of these AI-based techniques over traditional approaches can be attributed to their ability to identify and simulate the complex nonlinear relationships between mineral occurrence and evidence characteristics [19,24]. In particular, in recent years, the rapid development of DL methods—such as convolutional neural networks [25,26], deep autoencoder networks [26,27], recurrent neural networks [23,28], generative adversarial networks [19,29], transfer learning [30,31], reinforcement learning [32], and their various models [26,33,34]—has allowed them to obtain high-level features from input data and achieve encouraging ore prediction results [1,30]. However, as the number of input evidence features increases, it may lead to increased model complexity and an increase in the number of model parameters, making the model prone to overfitting [26]. Meanwhile, because data-driven MPMs usually ignore domain knowledge and experience, they have poor generalization ability and interpretation ability [26,35]. Therefore, data-driven MPMs are only suitable for brownfield areas with deep exploration, full mineral development, and rich data [9,11].

In addition, many scholars have used hybrid driving methods to integrate expert knowledge, such as fuzzy logic or fuzzy reasoning, into data-driven MPMs to predict mineral exploration targets [21,36]. These hybrid methods complement the expert knowledge in data-driven MPMs and compensate for the incomplete utilization of useful information in knowledge-driven or data-driven MPMs [10]. However, there are numerous uncertainties in these hybrid approaches due to the imbalance in geological exploration, multiple possible sources of minerals, and the diversity of physical and chemical information [31]. Meanwhile, because of the wide spatial range and numerous factors influencing the evolution of underground geological structures and mineral deposits, it is difficult to predict deep hidden mineral resources due to incomplete information on geological structures and their relationships and evolution. To address these challenges, some scholars used three-dimensional (3D) prospectivity mapping for mineralization prediction, which has demonstrated promising results [37]. However, due to incomplete geological structure information and the interrelationships between and evolution of deep buried ore deposits, the model parameters remain inadequately characterized [38,39]. Thus, some prospecting prediction methods that are effective for outcrops and shallow strata have proven difficult to apply to deeper deposits [5,11]. These difficulties include limited applicability and effectiveness of existing metallogenic series, metallogenic systems, metallogenic patterns, and prospecting prediction models, and limitations of the completeness, accuracy, and reliability of geological, geophysical, geochemical, and remote-sensing data [36,40,41].

Therefore, to address the limitations of these methods and utilize the results of the theoretical models without being constrained by their limitations, we adopted a technical

strategy that combines (i) model-based and model-free approaches, (ii) data-driven and model-driven approaches [3,9], and (iii) correlation relationships between ore deposits and other factors and their genetic relationships. At the same time, to reflect the value of data without being restricted by accuracy, it is necessary to adopt a methodology that combines (i) big data and small data, (ii) all data and random samples, and (iii) deep data mining and statistical analysis [41,42]. We developed a multi-scale progressive big data mineral prospectivity model (BDMPM) based on the fourth paradigm of scientific research [43,44] and big data, which uses spatiotemporal geological, geophysical, geochemical, and remote-sensing big data to conduct broad interdisciplinary integration, fusion, and deep mining to predict deep hidden solid mineral resources.

Songtao is located in the center of the NE-trending rift depression in the Tongren Paleorift. It is an area with the thickest Datangpo Formation sediments and the most enriched Mn mineralization. Four world-class super-large concealed Mn deposits have been proven here, making it the only concentrated area of super-large Nanhua Period Datangpo-type Mn ore (DMO) in the world [45,46]. Moreover, Songtao possesses great exploration depth and a unique metallogenic model, making it a “natural laboratory” and “touchstone” for comprehensively verifying the scientificity, effectiveness and robustness of the concealed deposit prediction model. Therefore, Songtao was used as a test case. We reviewed and summarized the actual mineral prospecting process in this area [47], adopted the idea of prospecting prediction based on geological anomalies [48], and followed a multi-scale progressive technical strategy from 1P to 5P (xP refers to various scales of mineralization potential prediction) [49]. We constructed key theoretical, methodological, and technical systems for progressive large- to small-scale prospecting prediction. Specifically, for 1P (probable ore-forming areas), we considered why Mn ore deposits are present in northeastern Guizhou by mining 1:500,000 province-scale gravity, aeromagnetic, and geological data; for 2P (permissive ore-finding areas), we addressed why Mn ore deposits are found in the Songtao ore-concentrated area by mining 1:200,000 regional basin-scale geochemical and geological data; for 3P (preferable ore-finding areas), we focused on the question of why Mn ore deposits occur in level-IV grabens [50] using a 1:50,000 mineral belt-scale prediction model; and for 4P (potential mineral resource areas) and 5P (prospective ore body areas), using 1:25,000 mineral-field-scale remote-sensing associated-mineral extraction and 1:5000 mineral deposit-scale borehole data, we resolved the distributions of individual Mn ore bodies.

2. Geological Background

The research area in northeastern Guizhou (27°59′–28°19′ N, 108°38′–109°23′ E) is adjacent to Hunan Province and Chongqing City and covers a total area of 3172.34 km² (Figure 1a). The northwestern part of the study area is the Upper Yangtze Paleocontinent, while its eastern side is the Xuefengshan uplift belt. The geological structures within the area are complex, with tectonic features trending in four directions: NNE, NE, NEE, and NW (Figure 1b). The DMO concentration area is located in the NE-trending fault-depression center of the Nanhua Rift, which formed during the breakup of the Rodinia supercontinent on the southeastern margin of the Yangtze Block. This area represents the most important Mn-bearing host horizon of the Neoproterozoic in South China [51,52]. The sedimentary facies, from the basin margin to the center, are successively divided into littoral, shelf, restricted lagoon, and deep-water fault-depression basin facies, constituting a complete transgressive–highstand–regressive sedimentary cycle.

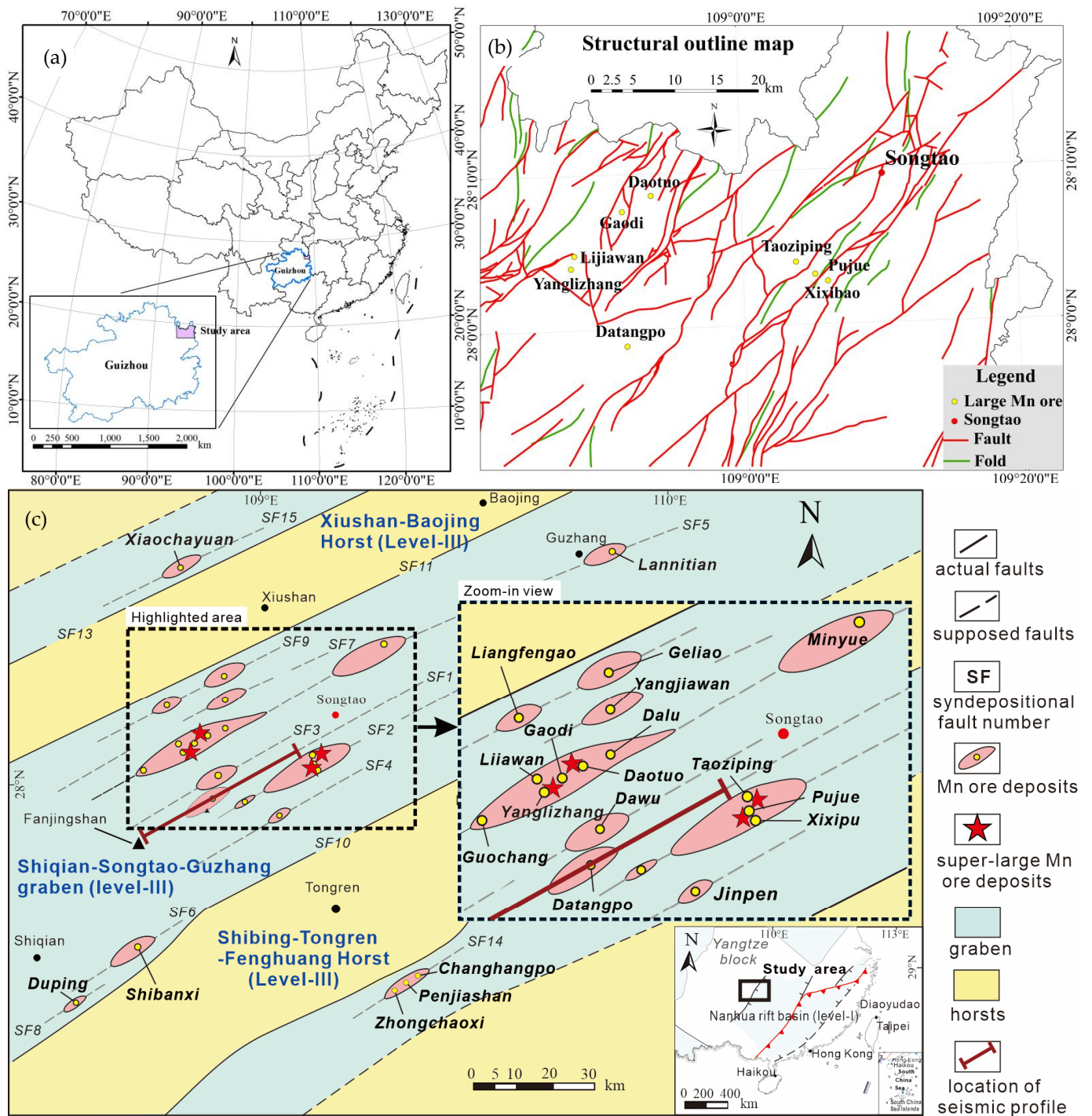


Figure 1. Geographical location and geological map of the study area: (a) geographical location, (b) structural outline map, and (c) spatial distribution of the DMO deposits.

Exposed strata are the Middle Paleoproterozoic Fanjingshan Group, Latest Paleoproterozoic Qingbaikou System, Nanhua System, Sinian System, and Early Paleozoic Cambrian System. The Nanhua System, which is the sedimentary cover of the Neoproterozoic Nanhua Rift formed after the breakup of the Rodinia Supercontinent [41], consists of, from oldest to youngest, the Liangjihe Formation, Tiesizao Formation, Datangpo Formation, and Nantuo Formation. Among these formations, the ore-containing interval of the Datangpo Formation exhibits a reverse-graded sequence with finer sediments at the base and coarser sediments upward, ranging in thickness from tens of meters to over two hundred meters, and thickening progressively from the basin margin toward the fault-depression center. The

ore-bearing succession can be subdivided into three successive sedimentary microfacies units: the basal fault-depression basin facies consists of deep black carbonaceous shale intercalated with thin-bedded tuffaceous claystone, with a total organic carbon (TOC) content exceeding 6%, deposited in a strongly reducing, restricted deep-water environment; the middle Mn-containing interval is characterized by interbedded rhodochrosite, manganocalcite, and Mn-bearing limestone, associated with hydrothermally derived Mn minerals and exhibiting typical exhalative-sedimentary structures; the upper restricted shelf facies comprises grayish-black thin-bedded silty shale, where the environment shifted from oxidizing to weakly reducing, accompanied by a rapid decrease in manganese content. From the fault-depression center toward the basin margin, the thickness of the Mn-containing interval decreases from more than ten meters to less than one meter, and the average ore grade drops sharply from over 25% to below 5%, demonstrating a pronounced “facies-controlled mineralization” characteristic [47,53,54]. The major Mn ore deposits in northeastern Guizhou are located within the Shiqian–Songtao–Guzhang level-III Mn-containing graben in the northwestern Nanhua Rift (Figure 1c) [50]. The ore-bearing layer is at the bottom of the Datangpo Formation [47] and is hosted within black carbonaceous shale.

After more than 60 years of constant exploration for, development of, and research on DMO Mn ore deposits in the research area, vast amounts of geological, geophysical, geochemical, remote-sensing, textual, and drilling data have accumulated [41]. New insights into the processes and controlling factors of metallogenesis have been produced, which have overturned the traditionally accepted international understanding of the genesis of Mn ore deposits. A new metallogenic mechanism for sedimentary exhalative deposits formed from deep Mn-containing fluids has been proposed and a corresponding prospecting model has been established [47]. Under the guidance of this prospecting model, a notable breakthrough has been achieved in the search for rhodochrosite deposits in northeastern Guizhou, resulting in the discovery of verified reserves containing more than 700 million tons [47,55]. The study area is suitable for utilizing previous prospecting experience of DMO Mn ore deposits to seek new breakthroughs in prospecting for hidden ores; in addition, the region is ideal for experimenting with a multi-scale progressive prospecting prediction method based on geological big data and the approach outlined above to achieve larger-scope prospecting breakthroughs for DMO deposits. To this end, the three-tier ore-controlling indicator system—fault intersection, fault-depression center, and favorable sedimentary facies belt—constructed in this study directly provides solid geological theoretical support for selecting variable and determining thresholds in the BOA-AdaBoost intelligent prospecting model.

3. Methods

3.1. Data Sources for Multi-Scale Prospecting Prediction

According to the principles of metallogenic system and geological anomalies, we listed the factors that are relevant to the metallogenic conditions and ore-controlling factors [55,56], and collected data pertaining to these factors to construct the prospecting prediction datasets (Table 1).

Using these data, we performed cross-boundary data convergence, assimilation, fusion [41], and deep mining to conduct mineral prospecting prediction and target delineation.

Table 1. Multi-source and multi-class heterogeneous prospecting prediction datasets.

Data Source	Category	Specific Name
Geology	Ore-bearing rock series and horizon	1:500,000, 1:200,000, 1:50,000, 1:25,000, and 1:5000-scale geological maps
		Stratigraphic units (group, segment)
	Lithology of the ore-bearing strata	Clastic rocks, carbonate rocks
	Faults	Type (normal fault, reverse fault), level (class I, II, III)
	Folds	Type (syncline, anticline, synclinorium, anticlinorium), grade (large, small)
	Distribution of mineral resources	Ore spots, ore deposits, ore bodies
	Occurrence mode of ore bodies	Main types (lamellar, quasi-lamellar, lenticular, veined) Minor types (lamellar, quasi-lamellar, lenticular, veined)
	Wall rock alteration	Alteration type, altered minerals, alteration intensity
	Associated minerals	Name, content
Geochemical exploration	Geochemical map	1:200,000
	Superposition of co-associated elements	Pb series, Cu series, Li series, Cd series, etc.
	Superposition of all elements	39 elements
Geophysical exploration	Gravity	Bouguer anomaly
		1:200,000
		Upward continuation
		2 km, 6 km, 10 km, 20 km
	Magnetism	Wavelet analysis
		Order 1, order 2, order 3, order 4
		Trend analysis
Remote sensing	Hyperspectral, multispectral	Order, trend, residue
		Aeromagnetic ΔT
		1:200,000
	Magnetism	Aeromagnetic ΔT polarization, aeromagnetic ΔT polarization vertical first derivative
		Trend analysis
		Order, trend, residue

3.2. Multi-Scale Data Mining and Prospecting Prediction Based on Big Data

A metallogenic system is a complex giant system, and minerogenesis is a nonlinear process that is the product of a geological anomaly. Prospecting prediction is the estimation or inference of metallogenic events that occurred in the past and their characteristics. In this study, we regarded metallogenic systems as being heterogeneous and considered the regional geological background and previous prospecting experience. First, multifactor data assimilation and fusion were performed on progressive scales based on big data, and the metallogenic conditions and ore-controlling factors of the DMO deposits were assessed through model-free mining of various types of geological anomalies. Then, a knowledge graph was established by reviewing the prospecting experiences associated with a series of Mn ore deposits. This graph was integrated with the results from big data mining to achieve combined model-based and model-free prediction. Finally, with the support of a 3D geological model, data-and-model-driven prospecting prediction was conducted using AI methods to identify specific target areas (Figure 2).

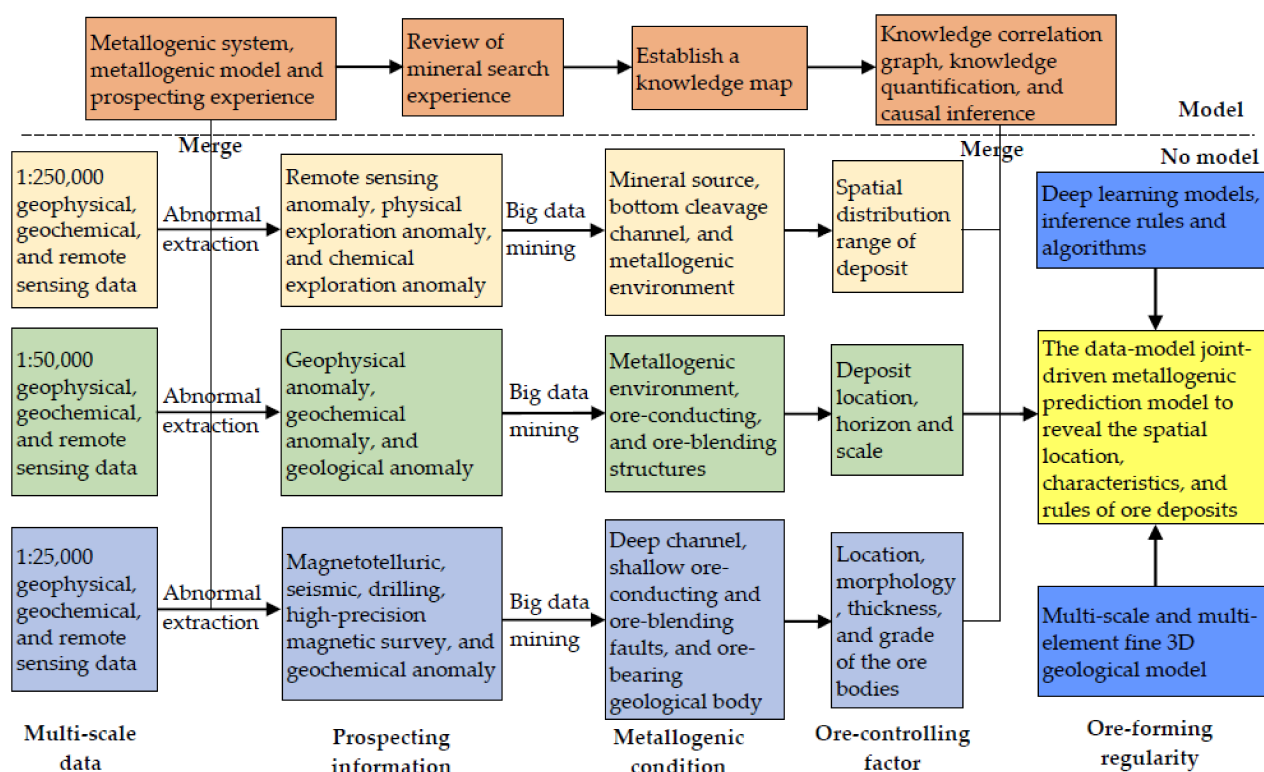


Figure 2. Schematic diagram of multi-scale progressive data mining and prospecting prediction methodology.

Different types of geological data differ regarding their granularity, breadth, and depth. These data represent an externalized presentation of the multi-scale, -directional, -dimensional, -level, and -modal characteristics of geological objects. Data mining for prospecting prediction involved analysis, coupling, and localization of geological anomalies [48,49]. Identification of anomalies progressed sequentially from ore-forming geological anomalies → exclusive geological anomalies → comprehensive geological anomalies → ore deposit geological anomalies → ore body geological anomalies. Prediction targets were also identified in sequence from probable ore-forming areas (1P) → permissive ore-finding areas (2P) → preferable ore-finding areas (3P) → potential mineral resource areas (4P) → prospective ore body areas (5P) (Figure 3). The working scale used and the level of detail of the information reflected therein also increased from 1:500,000 (1P, at the provincial scale) → 1:200,000 (2P, at the regional basin scale) → 1:50,000 (3P, at the mineral belt scale) → 1:25,000 (4P, at the mineral field scale) → 1:10,000 or 1:5000 (5P, at the mineral deposit scale) (Figure 3). It is noteworthy that end-to-end validation is required for the entire “1P-5P” framework. Only the regions where the previous P level outputs high mineralization anomalies or high probabilities are selected as inputs or spatial constraints for the next P level. Meanwhile, the outputs of each P level must be discussed with mineral exploration personnel to confirm the correctness of the results. In addition, whether each step is necessary depends on whether the data completeness at that P level can support the predictive computation.

As the amount of information increases, the scope of the prediction progressively narrows and the prediction target is brought into sharper focus. In other words, data mining systematically progresses from the macroscale to the microscale, from a region to a point, from shallow to deep, and from monomorphic to polymorphic in this process. There is also a shift from mining of single parameters to convergence, fusion, and mining of multi-source and multi-category heterogeneous data. To undertake prospecting prediction using geological spatiotemporal big data, it is necessary to carry out transboundary aggregation, connection, and systematic mining, and to closely integrate data mining technology

with knowledge gained from geological surveys and mineral exploration, geophysical–geochemical–remote-sensing technologies, and data management technology [56].

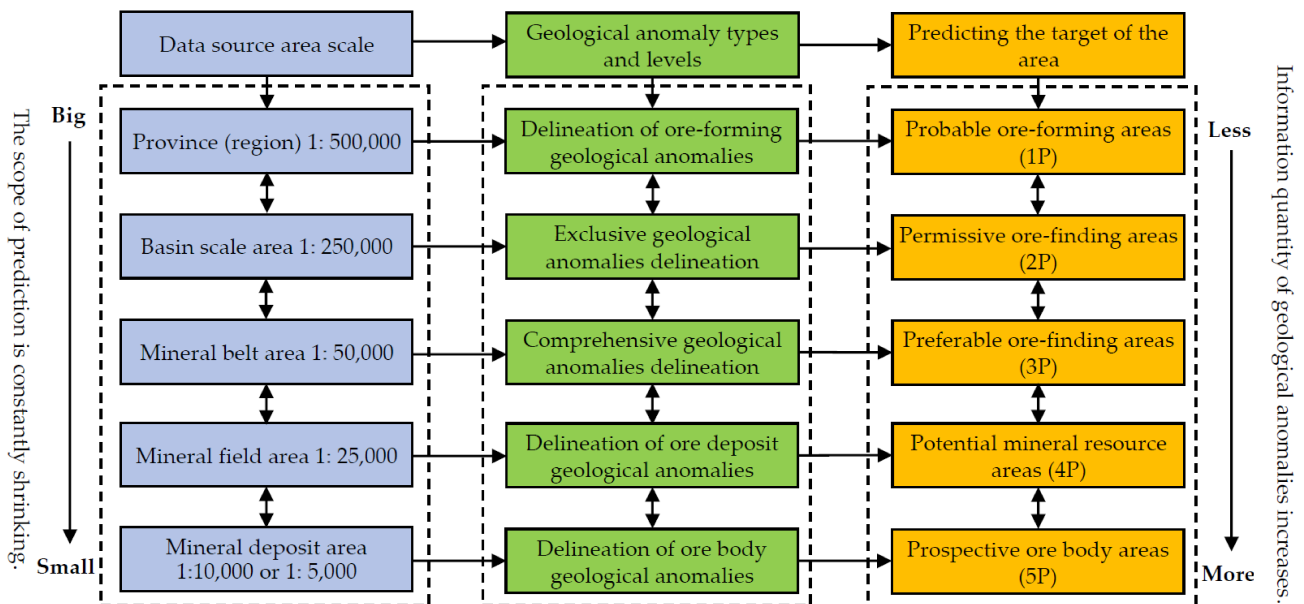


Figure 3. Different types of geological anomalies and corresponding grades of ore deposit prediction zones (modified from [55]).

3.3. Development of Multi-Scale Progressive BDPMP

According to the aforementioned prospecting prediction methodology, to predict ore prospects using big data, it is necessary to combine information on the regional geological background, survey data at various scales, and prediction goals and tasks to formulate the basic methodology (Figure 4).

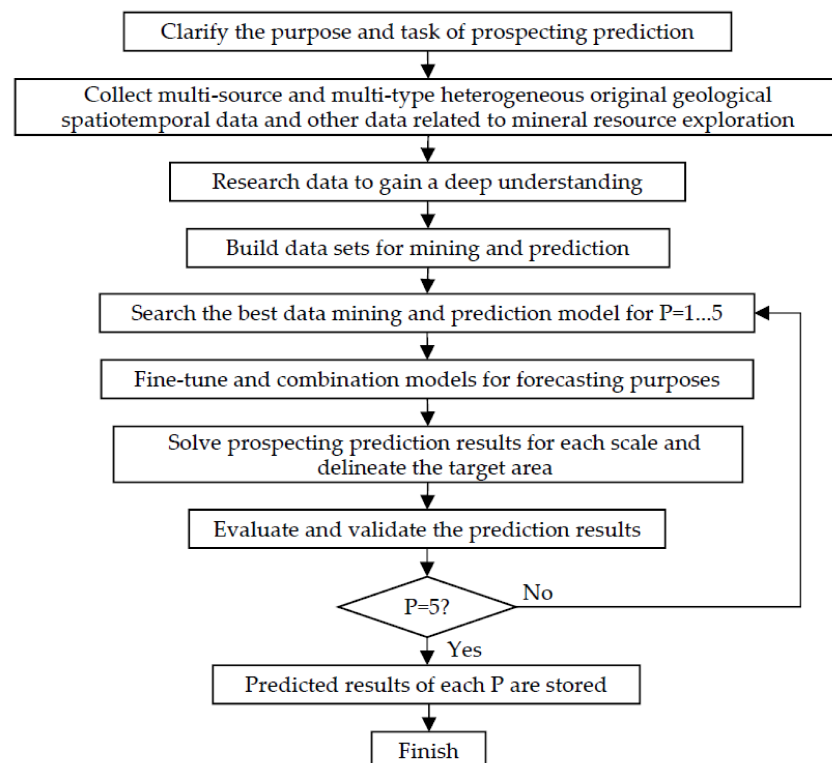


Figure 4. Procedure for multi-scale progressive prospecting prediction based on geological big data.

(1) The prime purpose of prospecting prediction is to identify target areas. Its task is to combine 3D geological models, AI methodologies, and data mining technologies to process the geological big data of the prediction area, and progressively identify the probable ore-forming, permissive ore-finding, preferable ore-finding, potential mineral resource, and prospective ore body areas, while gradually optimizing and then determining the primary target areas.

(2) Metallogenic data are collected; the exact data required will depend on the prediction scale and objective. Mining and prospecting prediction requires collection of data with super-coverage that is as close to the idea of geological big data as possible. Instead of focusing on a small amount of artificially extracted characteristic data, prediction using big data makes use of a wide range of multi-source and multi-type original geological spatiotemporal data and other data related to mineral resource exploration. Thus, the problems associated with using random samples do not arise because the datasets are used in their entirety [40].

(3) The data are analyzed to understand the data types and characteristics used in mining and prospecting prediction. This process involves the following steps: (i) obtaining an in-depth understanding of the data characteristics (e.g., statistical characteristics, data quality, and semantic characteristics); (ii) reviewing the historical prospecting experience and metallogenic patterns in the prediction area; (iii) assessing the fundamental geological characteristics of the prospective area from geological reports, the literature, and expert consultation; and (iv) analyzing the data to elucidate the metallogenic mechanisms, metallogenic conditions, and ore-controlling factors.

(4) The raw data are standardized and formalized [56] (ensuring that no information is lost), and a prediction data warehouse is constructed to improve the accessibility of the regional geological–geophysical–geochemical–remote-sensing data when constructing mining algorithms and prediction methods, and to support identification of potential data patterns.

(5) The optimal prediction model is explored and selected by analyzing the essence of the prospecting prediction problem (i.e., classification, clustering, and regression) and integrating the prediction objectives, data characteristics, and a priori knowledge graphs of the 1P objects.

(6) According to the actual situation and data characteristics, the prospecting prediction model is fine-tuned by adjusting the model parameters, quantifying the weights of the metallogenic elements, and establishing the model evaluation criteria (e.g., AUC-ROC, precision-recall curves, F1-score) and validation strategies (e.g., k-fold cross-validation, spatial cross-validation) to prevent overfitting. It is worth noting that the prediction results are not produced by a single model, but by combining multiple models for an optimal solution.

(7) The prediction results obtained using the 1P–5P approach are evaluated and a small amount of engineering validation is initiated. Then, with reference to the verification results and feedback from a field-based survey, the data are reorganized and the prospecting prediction models are adjusted and improved.

(8) After completing the prospecting prediction from 1P to 5P, all the prediction materials and results should be stored in relevant databases and maintained appropriately.

3.4. Adaptive Boosting (AdaBoost) Model and Its Hyperparameters

AdaBoost has been proven to be an effective and practical method for ensemble learning. By analyzing the subtle differences in each sample, it enhances overall classification accuracy through nonlinear mapping, which accumulates the product of weak classifiers and their weight coefficients.

Suppose the initial training set sample form is $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, with a total of m samples. The weight of the k – *th* weak learner updated for the training

set is set as $D(k) = (w_{k1}, w_{k2}, \dots, w_{km})$; $w_{1i} = \frac{1}{m}$; $i = 1, 2, \dots, m$. Then, the next set of weak learners $G_{k+1}x$ can be trained based on the dataset and weights. Mineral prospecting prediction primarily considers classification problems, and the weak learner is a classifier. Let the weighted error rate for the k – th round of classification to be defined as $e_k = \sum_{i=1}^m w_{ki} I(G_k(x_i) \neq y_i)$, where $I(G_k(x_i) \neq y_i)$ represents the mapping function between the predicted value and the actual value. When setting the weight coefficient of the classifier based on the error rate, the principle is that the larger the error rate, the smaller the weight coefficient; conversely, the smaller the error rate, the larger the weight coefficient. The weight coefficient of the k – th weak classifier G_kx is set as $\alpha_k = 1/2 \log(1 - e_k)/e_k$. Subsequently, the training set weights $D(k)$ are updated. By setting the weight of the k – th round for training set as $D(k) = (w_{k1}, w_{k2}, \dots, w_{km})$, the weight of each training sample in the $k + 1$ – th round can be obtained based on the weight coefficient of the classifier and the weight of the samples in the previous round. The formula is as follows:

$$w_{k+1,i} = \frac{w_{ki}}{Z_k} \exp(-\alpha_k y_i G_k(x_i)) \quad (1)$$

where w_{ki} represents the weight of sample i in the previous round, α_k represents the weight coefficient of the weak classifier in the previous round, y_i represents the true value, and $G_k(x_i)$ represents the predicted value. Z_k is the normalization factor, and its structure is

$$Z_k = \sum_{i=1}^m w_{ki} \exp(-\alpha_k y_i G_k(x_i)) \quad (2)$$

The normalized factor is used to calculate each sample's loss on the global scale. As shown in Equation (1), when $\alpha_k y_i G_k(x_i) < 0$, which indicates a classification error, the weight coefficient of this sample in the next iteration will significantly increase; conversely, if $\alpha_k y_i G_k(x_i) > 0$, it will decrease, which conforms to the basic setting. The sample update coefficient formula is derived from the optimization of the AdaBoost loss function.

Finally, weighted voting method is used for the final classification, as shown in Formula (3):

$$f(x) = \text{sign}\left(\sum_{k=1}^n \alpha_k G_k(x)\right) \quad (3)$$

AdaBoost is an MPM constructed by updating the weights of each sample in the original dataset and then forming a set of classifiers. The weak learners in it adopt a decision tree (DT). The algorithm flowchart is shown in Figure 5, and the specific steps are presented in Table 2.

Table 2. Steps of AdaBoost algorithm.

Step	Step Description
1	Standardize and normalize the dataset, and allocate the training, test, and validation sets at a ratio of 7:2:1.
2	Assign the same weight to each sample, train the first DT ₁ (the initial weight of the first classifier is set to 1), let DT ₁ classify each sample x_i in the training set to obtain the predicted value $G_1(x_i)$, and calculate the error rate e_1 according to the formula $\sum_{i=1}^m w_{ki} I(G_k(x_i) \neq y_i)$.
3	The algorithm iterates through rounds using the error rate e_1 . The classifier's own weight α_1 for this round is updated based on the formula $\alpha_k = 1/2 \log(1 - e_k)/e_k$. Then, the weights of each sample for the next round $w_{2,i}$ are calculated according to Formulas (1) and (2), and by integrating e_1 , α_1 and $G_1(x_i)$. Subsequently, the second classifier DT ₂ is trained based on the sample set and sample weights, and this process continues iteratively.
4	Repeat step 3 continuously to construct 15 weak classifiers (15 rounds of iterations) and 15 classifier weights α_k . These are then combined into a strong classifier using the cumulative voting method (Formula (3)).

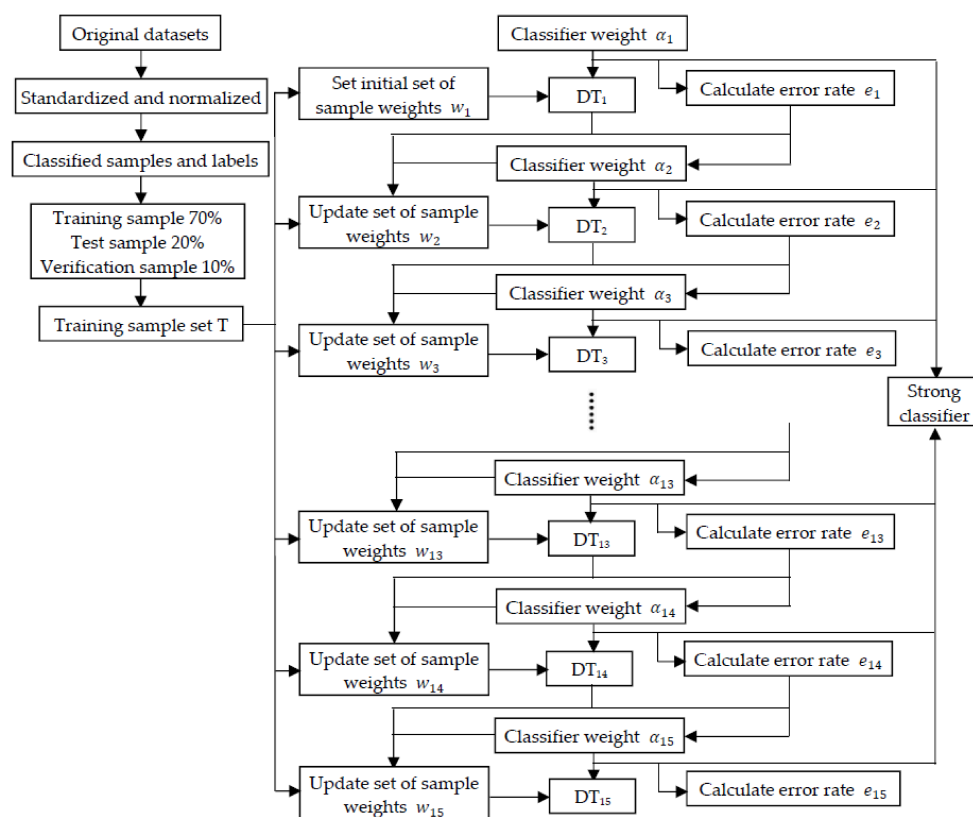


Figure 5. Flowchart of the AdaBoost algorithm.

To avoid the classification bias caused by imbalanced data, the mean square error loss function is used to correct the cost-sensitive loss function of this shallow neural network:

$$Loss = \frac{1}{2} \sum_{i \in Examples} ((t_i - p_i) \cdot cost(n, n))^2, n \in 0, 1 \quad (4)$$

where $cost(n, n)$ represents the weight that should be assigned to a prediction situation in the confusion matrix, and t_i and p_i are the true classification and predicted classification of the sample, respectively.

The AdaBoost model has three types of parameters: general, booster, and learning [10]. According to the characteristics of the Mn ore samples, a decision tree (DT) is selected as the base_estimator of the model, and boosting is carried out based on the DT classifier. General parameters such as n_estimators, which is the number of iterations of the base classifier, are used to adjust the model's complexity and prediction accuracy. Excessive values may lead to overfitting, while insufficient values may result in underfitting. Booster parameters are the most important and include max_depth, min_samples_split, and min_samples_leaf [10]. Max_depth is the depth of the DT, which limits its complexity. Min_samples_split is the minimum number of samples required to split a node in the DT. Min_samples_leaf is the minimum number of samples required to be at a leaf node in the DT. These three parameters are used to avoid overfitting in the model. Learning_rate, which is the weight reduction coefficient of each weak classifier, balances stability and training time while controlling learning speed and generalization ability. Random_state sets up a random seed that is usually the same in all tests to ensure the repeatability of results.

By adjusting these parameters, the stability and repeatability of the AdaBoost model can be enhanced, enabling it to better adapt to specific datasets and prediction tasks. Meanwhile, techniques such as cross-validation can be employed to determine the optimal parameter combinations.

3.5. Bayesian Optimization Algorithm (BOA)

Improper initialization parameters can reduce the performance of ensemble learning models [10]. To address this, hyperparameters can be optimized to enhance the accuracy, robustness, and reliability of AI models [57]. BOA can quickly obtain the optimal values and is widely used to determine the optimal hyperparameter values of models [58,59]. In this study, BOA was employed to optimize the hyperparameters of the AdaBoost model. The probability model of BOA uses Gaussian processes (GPs) to calculate the predictive distribution of the objective, as detailed below.

Suppose the optimized black-box objective function $P(x)$ is drawn as a random sample from a GP, expressed as $P(x) \sim GP(\mu(x), k(x, x'))$, where $\mu(x)$ represents the mean function and $k(x, x')$ refers to a covariance function. For a multi-output case, it is extended to a multi-task GP: $P(x) \sim GP(\mu(x), K(x, x'))$, where $K(x, x')$ is formulated through the coreization kernel, integrating the covariance structure of all feature vectors:

$$P(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |cov|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T cov(x - \mu)^{-1}\right) \quad (5)$$

where μ (mean) and cov (covariance) are defined as follows:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

$$cov = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (7)$$

As the number of samples grows incrementally, BOA is able to approximate the posterior distribution of the objective function, leading to the identification of the optimal value and ultimately enabling hyperparameter tuning for the model.

3.6. Data Processing and Sample Selection

The study region was divided into a 30 m $\times$ 30 m grid, with a total of 3,524,826 grid cells in Songtao. Meanwhile, a 23-dimensional dataset for Mn ore from geological, geophysical, geochemical, remote-sensing data in Songtao was collected, which included stratum, fold, fault, ΔT aeromagnetic isoline, ΔT polar isoline, ΔT polar vertical first derivative isoline, gravity continuation at 2, 6, 10, and 20 km, chlorite alteration anomaly, manganese oxide alteration anomaly, and As, Co, Cr, Cu, Mn, Ni, Pb, Zn, Al_2O_3 , CaO, Fe_2O_3 , MgO, and SiO_2 data [41]. These factors were then normalized and classified into different grades based on their level of influence on the formation of Mn ore deposits [56]. Finally, all these data were used to generate a database for the MPM for Mn ore deposits in Songtao. In the database, 647 positive samples from ore-bearing boreholes and 647 grid points selected as negative samples far away from these positive samples were used to construct the training, test, and validation sets at a ratio of 7:2:1. Among them, the selection of negative samples was based on the global Moran's I statistics of geoscientific variables in the Songtao Mn ore concentration area, with the spatial autocorrelation critical distance of 3.2 km taken as the core threshold. At the same time, by overlaying metallogenic factor constraints and a 1 km grid stratified sampling rule, the spatial autocorrelation distance of the final sample set was controlled within 0.8 km. This fundamentally avoided spatial autocorrelation bias and prevented overestimation of model performance caused by arbitrary empirical sampling.

4. Multi-Scale Progressive Prediction Results for Hidden Mn Ore Deposits

We carried out 1P–5P [49] multi-scale progressive prospecting prediction using the method outlined above based on geological, geophysical, geochemical, and remote-sensing data for the Songtao Mn-ore-concentrated area of northeastern Guizhou [41]. The prediction

involved the complete step-by-step delineation of a target area from identification of ore-forming geological anomalies → exclusive geological anomalies → comprehensive geological anomalies → ore deposit geological anomalies → ore body geological anomalies. At the same time, the prediction method and key technical system were established from probable ore-forming areas (1P) → permissive ore-finding areas (2P) → preferable ore-finding areas (3P) → potential mineral resource areas (4P) → prospective ore body areas (5P) [49].

4.1. Prediction of Probable Ore-Forming Areas (1P)

The main purpose of identifying probable ore-forming areas (1P) is to establish the geological background and regional conditions, and to determine whether large-scale enrichment of mineral matter is present. In this study, this constituted identifying the geological conditions in the Songtao Mn-ore-concentrated area. Through processing and mining of 1:200,000 gravity, aeromagnetic, and geochemical anomaly data, we detected a hidden NNE-trending Yanshanian Songtao–Rongjiang foreland depression belt in eastern Guizhou (Figure 6). The DMO Mn ore concentration in Songtao, northeastern Guizhou, occurs in the northern part of this foreland depression belt (Figure 6). Existing studies have shown that large or super-large Mn deposits (Daotuo, Gaodi, Lijiawan, Yanglizhang, Datangpo, Xixibao, Pujue, and Taoziping) have been discovered in the Songtao region of northeastern Guizhou [47]. The Mn-containing rock series in the middle and lower parts of the Nanhua Rift sedimentary cover have not been eroded and destroyed by later geological processes because they were protected by the foreland depression. Therefore, regarding the metallogenic system, the Nanhua System in the Yanshanian Songtao–Rongjiang foreland depression belt should be considered to be a probable ore-forming area (1P) (Figure 6).

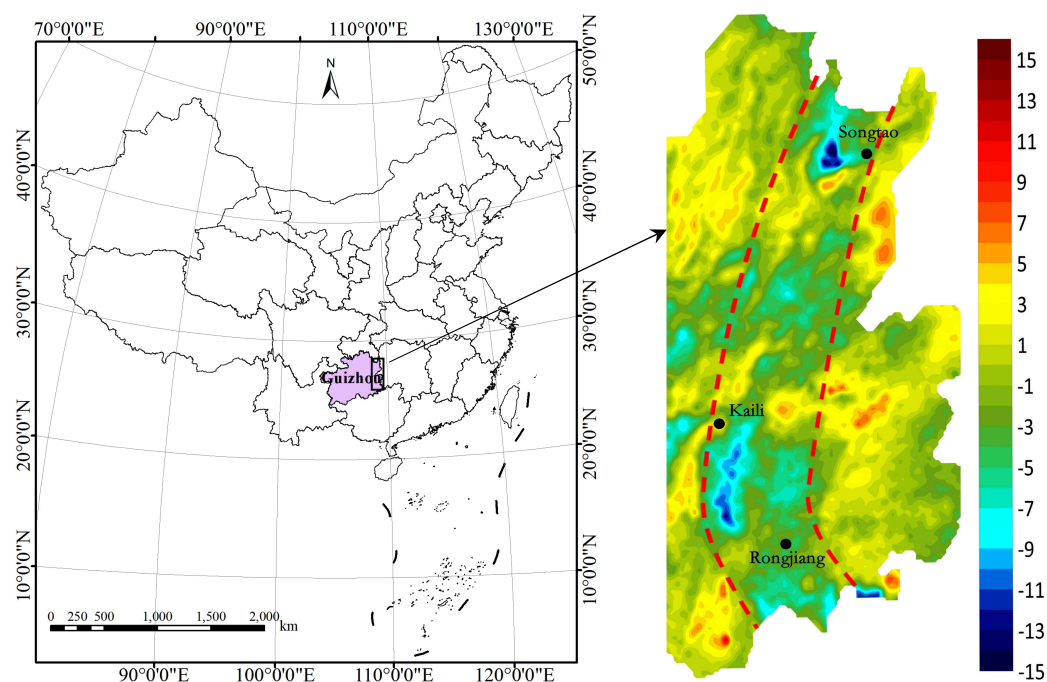


Figure 6. The Songtao–Rongjiang foreland depression belt in eastern Guizhou identified by gravity, aeromagnetic, and geochemical data (the color scale on the right indicates the depression index).

4.2. Prediction of Permissive Ore-Finding Areas (2P)

The main challenge in predicting permissive ore-finding areas (2P) is determining the direction of mineral concentration and migration; in this study, this involved identifying the factors controlling the location of the Songtao ore-concentrated area. We used aeromag-

netic ΔT anomaly data to visualize the tectonic framework of the upper crust; Bouguer gravity anomaly data to obtain information on the tectonic framework of the middle–lower crust; and remote sensing, outcrop, regional geological survey, mineral exploration, and geochemical exploration data to assess the near-surface tectonic conditions. From all these data, we identified a NW-trending ancient rift, called the Tongren Rift, that was formed during deposition of the Fanjingshan Group. The existence of this rift has been confirmed by reflection seismic detection (Figure 7a) [47]. It is hypothesized that during the formation of the ENE-trending Nanhua Rift, the Shiqian–Songtao–Guzhang level-III graben in northwestern Guizhou was superimposed on the NW-trending Tongren Rift (Figure 7b) [47]. This interaction formed a large-scale vertical tectonic fracture zone throughout the depth of the rift that resulted in a welling up of a deep Mn-containing fluid diapir to the basement of the Nanhua Rift, causing discharge and deposition mineralization. The areas where the ENE-trending Nanhua Rift secondary tectonic unit and the NW-trending ancient rift intersect represent permissive ore-finding areas (2P) for Mn ore deposits in northeastern Guizhou [50]. The results are highly consistent with the existing research findings and the known Mn ore deposits in the Songtao region of northeastern Guizhou [47].

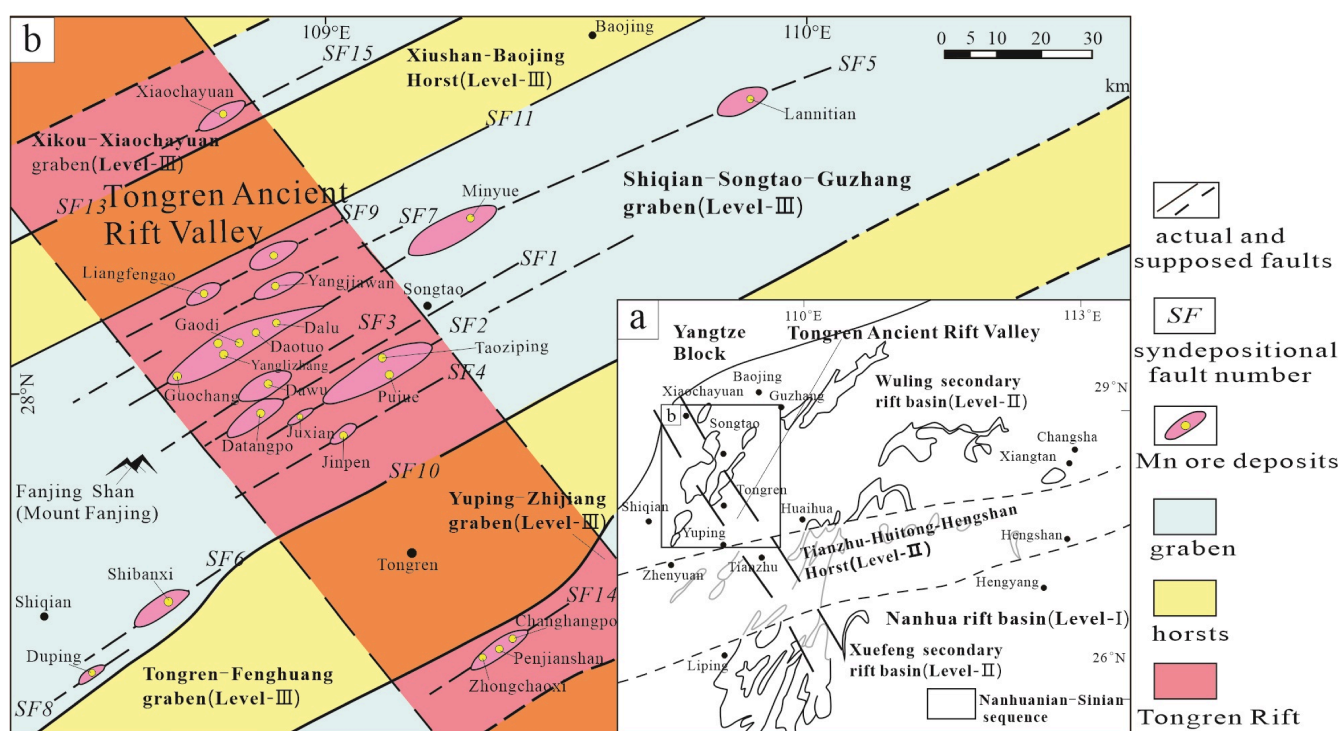


Figure 7. The NW-trending Tongren Rift identified from geological, gravity, aeromagnetic, and seismic data (from [47]). (a) Nanhua rift basin, (b) Tongren Rift and distributions of DMO deposits.

4.3. Prediction of Preferable Ore-Finding Areas (3P)

Many new types of Mn ore deposits in northeastern Guizhou are the products of discharge and deposition mineralization from deep-sourced Mn-containing fluids in the marine environment of the Nanhua Rift [47]. Preferable ore-finding areas (3P) are typically situated around Mn-containing gas–liquid exhalative centers. The key steps of the prediction methodology are as follows. First, the geological, geophysical, geochemical, and remote-sensing parameters related to rich ore bodies (reflecting the locations of the exhalative centers) are determined by reviewing the prospecting process and assessments of multiple known deposits. Then, prospecting prediction models based on AI are established to identify geological anomalies and to identify potential gas–liquid exhalative centers. Finally, preferable ore-finding areas (3P) are identified.

4.3.1. Results of Model Hyperparameter Optimization

The purpose of hyperparameter optimization is to ensure that the AI algorithms achieve optimal performance. BOA is used to optimize the hyperparameters of the AdaBoost model to reduce the predefined loss function and improve the prediction accuracy of the model. During the BOA-optimized ensemble learning process, the input parameters encompass all settings for the AdaBoost model, while the output is the results after 5-fold cross-validation based on the relationship between calculation time and deviation (Table 3).

Table 3. Hyperparameter optimization process and BOA-AdaBoost model results.

Step	Accuracy	Max_Depth	Learning_Rate	Min_Samples_Split	Min_Samples_Leaf	N_Estimators
1	0.913	5	0.325	19	16	19
2	0.901	6	0.961	8	19	45
3	0.907	7	0.817	6	15	141
4	0.921	8	0.814	15	4	119
5	0.916	14	0.728	13	11	172
6	0.897	4	0.296	7	14	200
7	0.918	10	0.878	9	3	126
8	0.908	3	0.859	10	12	78
9	0.903	12	0.427	17	10	86
10	0.905	13	0.266	14	17	120
11	0.914	2	0.359	3	6	123
12	0.936	9	0.576	11	5	176
13	0.895	1	0.496	12	7	151
14	0.926	8	0.378	17	18	18
15	0.915	15	0.362	16	20	196
16	0.917	11	0.572	18	2	15
Best	0.936	9	0.576	11	5	176

Table 3 shows that the parameters obtained through BOA optimization may not be the best. Therefore, it is necessary to select a better set of hyperparameters from the search results. Details are as follows: first, candidate parameter pools with the top 15% cross-validation accuracy are selected from all iterative historical results; then, screening is performed through generalization error validation on an independent test set, stability examination of weak classifier weight distributions, and spatial continuity assessment of mineralization prediction; finally, based on Occam's razor criterion, a lightweight parameter combination that balances predictive accuracy and generalization capability is selected. The optimal parameters for the AdaBoost model were [9, 0.576, 11, 5, 176]. Using these parameters, the AUC of the BOA-AdaBoost model reached 0.967, which is higher than the 0.939 achieved by the AdaBoost model. The loss curves and validation accuracy of the AdaBoost model before and after hyperparameter optimization are shown in Figures 8 and 9, respectively.

Figure 8 shows that during the training of the MPM, the mean squared error (MSE) of the BOA-AdaBoost model decreased as the number of iterations increased, which is consistent with the expected behavior of model convergence. Specifically, the MSE of the optimized BOA-AdaBoost model was significantly lower than that of the unoptimized AdaBoost model ($p < 0.05$, as determined by t -test), indicating that Bayesian optimization can effectively adjust parameters and enhance generalization capability. The final MSE of the BOA-AdaBoost model quickly converged to a minimum value (0.1), which is attributed to the weighted sample update mechanism of the AdaBoost model and the strong adaptability of the learning_rate parameter after Bayesian optimization.

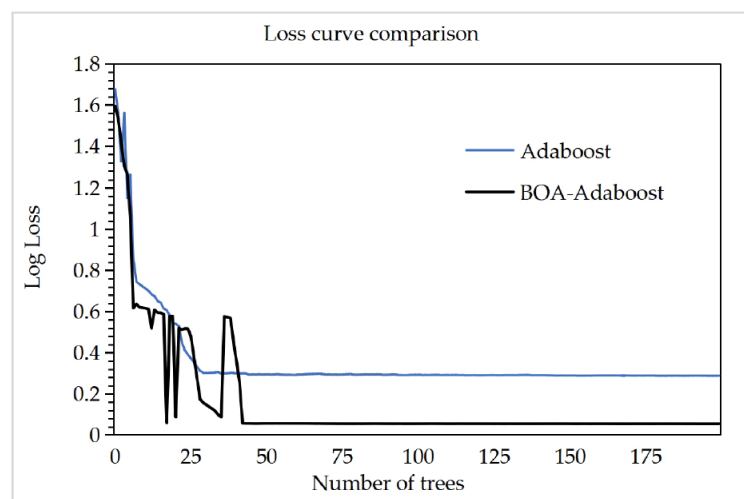


Figure 8. Loss function plots for AdaBoost and BOA-AdaBoost models.

Figure 9 shows that the accuracy of BOA-AdaBoost consistently outperforms that of AdaBoost throughout the entire iteration. In terms of convergence speed, BOA-AdaBoost requires only approximately 25 estimators to essentially converge, attaining a stable high-accuracy. In contrast, AdaBoost needs nearly 70 iterations to gradually converge. Regarding stability, the AdaBoost exhibits pronounced sawtooth-like oscillations throughout the entire process, with the minimum accuracy declining to 0.88. Conversely, the BOA-AdaBoost shows minimal overall fluctuation, experiencing only a slight upward peak around the 175th iteration, while maintaining accuracy consistently above 0.92 throughout, thereby demonstrating superior robustness. This indicates that BOA is an effective optimization for the AdaBoost model. Figure 9 also shows that the optimal parameters obtained through BOA are consistent with our previous experience: the larger the estimated $n_{\text{estimators}}$ value, the better the accuracy of the AdaBoost model; conversely, if $n_{\text{estimators}}$ reaches a certain high level, the accuracy of the AdaBoost model will fluctuate. Additionally, larger $n_{\text{estimators}}$ values necessitate more computing power and memory consumption. Therefore, we made a trade-off between training time and model effectiveness, and the BOA-AdaBoost model achieved its best performance when the value of $n_{\text{estimators}}$ was set to 176. Furthermore, the parameter optimization results indicate that the BOA-AdaBoost model can achieve higher accuracy by adjusting max_depth to 9.

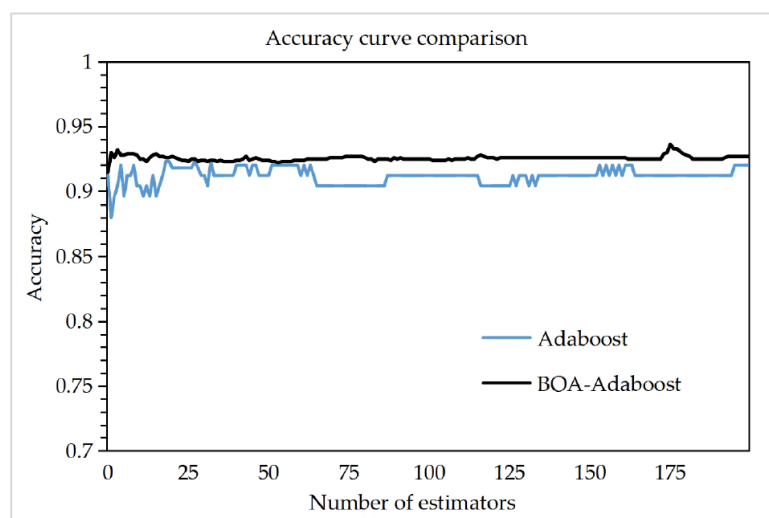


Figure 9. Validation accuracy plots for AdaBoost and BOA-AdaBoost models.

In summary, the AdaBoost model employs a Boosting algorithm, which reduces overall bias by iteratively adjusting sample weights and relying on the collaborative effect of weak learners with DTs as its base weak classifiers, which can effectively control variance. Understanding the trade-off between bias and variance is helpful for improving the fitting effect in Mn ore prediction and obtaining a more accurate MPM. Therefore, the hyperparameter optimization results of this study can provide practical guidance for ensemble learning modeling in similar scenarios.

4.3.2. Performance Metrics of AdaBoost and BOA-AdaBoost Models

Precision evaluation and validation analysis are numerical parameters for assessing the performance of ensemble learning models to prove the effectiveness and scientific validity of the methods used. This study adopted six indicators—namely, accuracy, precision, recall, F1 score, kappa, and AUC value—which are widely used to evaluate the performance of MPMs, to assess the performance of the AdaBoost and BOA-AdaBoost models (Table 4).

Table 4. Comparison of performance of AdaBoost and BOA-AdaBoost models.

Model	Accuracy	Precision	Recall	F1 Score	Kappa	AUC
AdaBoost	0.923	0.931	0.911	0.913	0.827	0.939
BOA-AdaBoost	0.936	0.948	0.925	0.928	0.862	0.967

The six indicators in Table 4 show that the accuracy of the BOA-AdaBoost model is 0.936, the precision is 0.948, the recall is 0.925, the F1 score is 0.928, the kappa is 0.862, and the AUC is 0.967, all of which are superior to those of the AdaBoost model. This indicates that the optimization of the AdaBoost model by BOA is effective and proves that BOA is a powerful optimization tool. It also indicates that the optimized BOA-AdaBoost model has higher accuracy and precision, better classification discrimination ability, and higher sensitivity. Additionally, Table 4 shows that the BOA-AdaBoost model achieved a better balance between accuracy and recall, along with improved overall consistency and classification performance.

The six indicators in Table 4 also demonstrate that the BOA-AdaBoost and AdaBoost ensemble learning models exhibit excellent predictive capabilities. This is because both models dynamically adjust the weights of the next DT based on their previous performance. Meanwhile, a regularization term is incorporated into the cost function to regulate the complexity of the model and reduce its variance to handle overfitting. Additionally, both models utilize multi-core CPUs to train weak learners in parallel, completing the parallel training of base classifiers. Furthermore, the BOA-AdaBoost model reduces the computational cost of hyperparameter search through surrogate models, accelerating Bayesian optimization and significantly improving model accuracy and running speed [60].

4.3.3. Prediction of Preferable Ore-Finding Areas (3P) Based on BOA-AdaBoost Model

A 23-dimensional dataset encompassing geological, geophysical, geochemical, and remote-sensing factors for Mn ore occurrence in Songtao were input into the optimized BOA-AdaBoost model to obtain the corresponding comprehensive anomaly values. Each point's predicted value was represented by a floating-point value ranging from 0 to 1. The larger the comprehensive anomaly value, the higher the probability of Mn mineralization. The comprehensive anomalies were divided into five levels: extremely low (0–0.2), low (0.2–0.4), moderate (0.4–0.6), high (0.6–0.8), and extremely high (0.8–1.0). Anomalies with values >0.6 can be regarded as preferable ore-finding areas (3P) for Mn ore deposit prospecting in northeastern Guizhou (Figure 10). The known large and super-large Mn

deposits (Daotuo, Gaodi, Lijiawan, Yanglizhang, Datangpo, Xixibao, Pujue, and Taoziping) are all located in the high metallogenic areas [47,56].

Figure 10 indicates that due to the influence of multiple episodes of tectonic activity, the comprehensive anomalies in the study area not only bear the traces of the NW-trending Tongren Rift in the basement of the Fanjingshan Group that formed in the pre-mineralization period, but also show imprints of the ENE-trending Nanhua Rift secondary graben and horst structure that formed during the mineralization period, as well as reflections of the NNE-trending Yanshanian tectonic features that formed after the mineralization period [47]. Figure 10 also indicates that the spatial distribution of these comprehensive anomalies is highly consistent with the probability of Mn mineralization and is influenced by factors such as thickness of overburden strata, lithology, and degree of destruction. In areas with a large overburden thickness (i.e., synclinal areas) where late-stage faults are not well-developed, comprehensive anomalies are weaker and the probabilities for mineralization are low, whereas in areas with a small overburden thickness (i.e., anticlinal areas) where late-stage faults are well-developed, the comprehensive anomalies are stronger and the probabilities for mineralization are high.

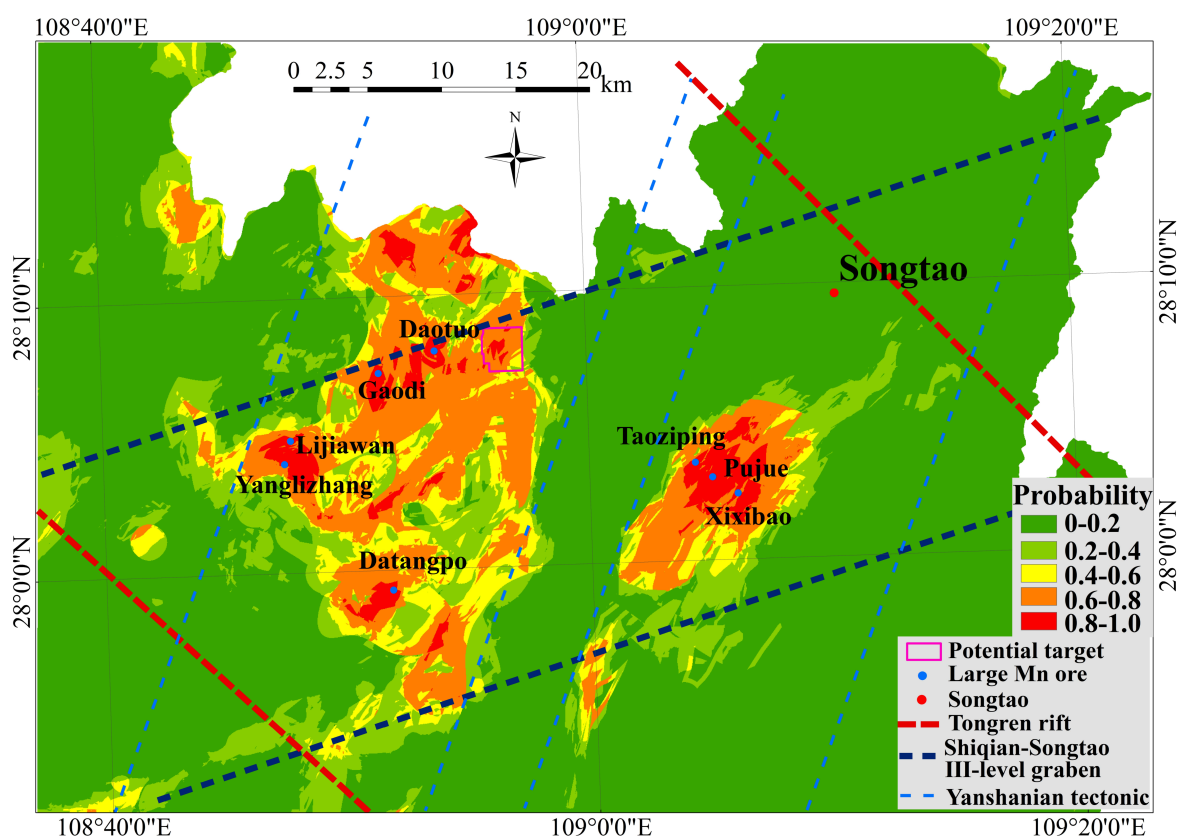


Figure 10. Distribution of comprehensive anomalies and delineation of 3P (>0.6) zones based on geological, geophysical, geochemical, and remote-sensing data on Mn ore deposits in northeastern Guizhou.

Furthermore, to better understand the workflow of the 1P-5P progressive mineral prospecting prediction, using the same Songtao dataset, a quantitative spatial comparison analysis between the extremely high-probability areas predicted by BOA-AdaBoost 3P and those predicted by the previously published DCNN (2024) [56] and Geo-WGAN (2025) [41] models was conducted (Table 5). The spatial overlay statistics show that the shared-area percentages are only 15.79% and 42.7%, respectively. In addition, the drilling success point mentioned in the paper (5.47 m, grade 18.66%, 23.08 million tonnes) was predicted

by both the 2025 Geo-WGAN model and the method proposed in this paper, indicating that both approaches are effective for predicting deep hidden Mn deposits, and further demonstrating the effectiveness of the 1P–5P prediction methodology system proposed in this paper.

Table 5. Comparison of shared-area percentages among different models (DCNN, Geo-WGAN, and BOA-AdaBoost).

Comparative Combination	Areas (km ²)	Shared-Area with BOA-AdaBoost 3P (km ²)	Overlap Ratio (Relative to BOA-AdaBoost 3P Area)	Verify Whether the Identified Ore Deposit Lies in the Overlapping Area
2024 DCNN extremely high-probability areas	73.47	8.72	15.79%	No
2025 Geo-WGAN extremely high-probability areas	29.15	23.58	42.7%	Yes
BOA-AdaBoost 3P extremely high-probability areas	55.22	55.22	100%	Yes

4.4. Prediction of Potential Mineral Resource Areas (4P)

Prediction of potential mineral resource areas (4P) is based on prediction of preferable ore-finding areas (3P). This work can draw on the techniques of conventional prospecting prediction, e.g., using remote-sensing information on alteration minerals and specific symbiotic, associated, and secondary minerals. This process involves five key steps: First, the ASTER images were preprocessed, which included vegetation disturbance removal, atmospheric correction, and de-crosstalk processing. Second, based on the spectral curve characteristics of pyrite (the main mineral associated with Mn mineralization), secondary Mn oxides, and chlorite, the mineral components were screened by principal component analysis (PCA) [61]. Third, the parameters of the support vector machine (SVM) were optimized using the ant colony algorithm (ACA) [61]. Fourth, remote-sensing anomaly information of Mn mineralization in the study area was extracted and combined to obtain spectral anomaly information for the symbiosis, associated, and secondary minerals of Mn mineralization. Finally, the combined anomaly areas were delineated based on the strong anomaly information and superimposed with the known deposits for analysis. There was a nearly 100% overlap between the results and known deposits [61] (Figure 11), indicating that identification of Mn ore deposits using the proposed method is accurate and reliable. This is because the training and validation sets were stratified according to the spatial blocks of ore-forming belts, ensuring that spatially adjacent ore occurrences do not appear simultaneously in both training and testing samples, thereby breaking the transmission pathway of spatial autocorrelation at the sample level. The 100% overlap with known ore occurrences achieved here was essentially a complete recapitulation of the key remote-sensing metallogenic indicators for Mn deposits by the model, rather than mechanical memorization of training samples. Therefore, the spectral anomaly areas of Mn symbiosis and associated (secondary) minerals can be considered potential mineral resource areas (4P).

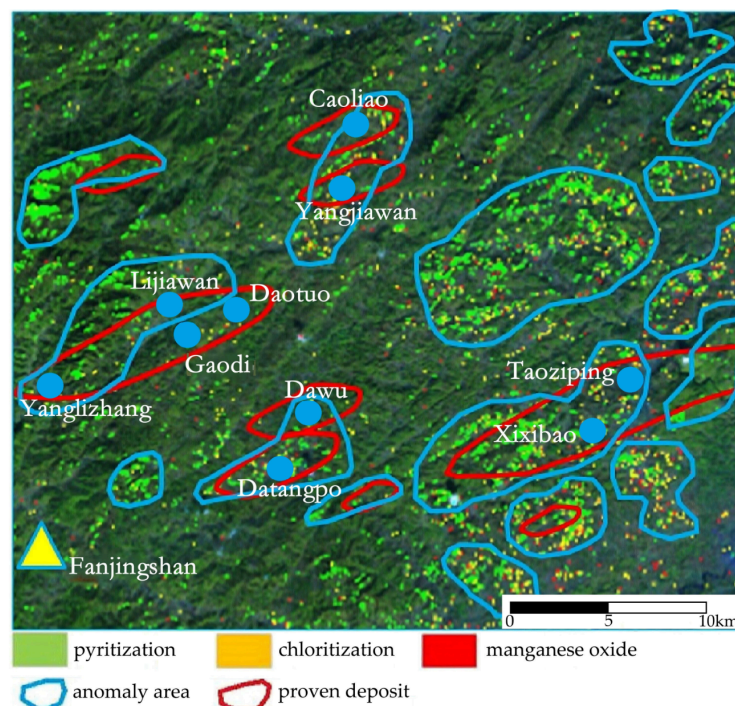


Figure 11. Comparison of combined anomaly areas with known deposits (from [61]).

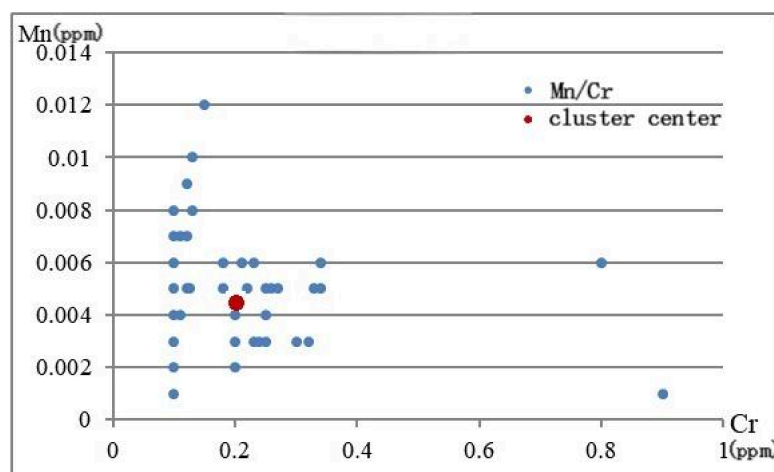
4.5. Prediction of Prospective Ore Body Areas (5P)

After prediction of potential mineral resource areas (4P), prediction of the prospective ore body areas (5P) can be performed. The main objective of this step is to obtain independent, clear, and reliable information that can directly reflect the existence of an ore body. In a previous study of multiple level-IV graben within the level-III graben of the Nanhua Rift System, the average ratio of Mn to Cr (Mn/Cr) of carbonaceous shale at the bottom of the first segment of the Datangpo Formation in the main ore-bearing section was found to be approximately 40; this value can be used to infer the presence of deep large or super-large hidden Mn ore deposits [47]. This geochemical indicator can be applied to predict deep prospective ore bodies, and even to determine whether there are level-IV grabens below the surface. This indicator was discovered after long-term exploration without any reference materials. To simplify and accelerate this exploration process, it is possible to automatically search for similar ore-indicating parameters based on big data methods and improve their universality. In this study, we used whole-ore chemical analysis data to review and track mining experiments.

Using all sample test data in the study area (Table 6), (1) correlation analysis was applied to trawl through all the geochemical exploration data in the region and search for symbiotic and associated elements; (2) multiple ratio analysis was adopted to construct element symbiotic combination modes; (3) informativeness analysis was carried out to determine the relationship between hidden deposits and particular geochemical indicators; (4) the analytic hierarchy process was employed to select prospecting models for hidden Mn ore deposits in the study area; and (5) various classification methods and typical AI algorithms were applied to assess the association relationship between the Mn/Cr ratio in Mn-containing rock series and the occurrence of hidden ore deposits and ore bodies (Figure 12). After completing the preprocessing of all borehole and outcrop test data, the relationship between the Mn/Cr ratio and the occurrence of hidden Mn ore deposits was investigated using various AI methods such as CNN, evolutionary algorithms, and random forest. The results confirmed that this manually obtained geochemical indicator has significant practical application value in the search for Mn ore deposits [47].

Table 6. Mn and Cr contents at different depths within the borehole, and the overall Mn/Cr ratio of the borehole.

Sample Location	Lithology	Stratum	Mn (ppm)	Cr (ppm)	Mn/Cr
YLZ-ZK-1	Mn-carbon shale	NhL ₂	142,500	15	/
YLZ-ZK-1	Carbon sandstone	NhL ₂	5750	33	/
YLZ-ZK-1	Carbonaceous shale	Nhd ₁	3250	35	/
YLZ-ZK-1	Carbonaceous shale	Nhd ₁	3000	43	/
YLZ-ZK-1	Carbonic clay	Nhd ₁	2100	38	/
YLZ-ZK-1	Carbonaceous shale	Nhd ₁	2000	40	/
YLZ-ZK-1	Carbonaceous shale	Nhd ₁	1550	45	/
YLZ-ZK-1	Carbonaceous shale	Nhd ₁	1500	30	/
YLZ-ZK-1	Carbonaceous shale	Nhd ₁	1050	43	/
YLZ-ZK-1	Gray clay rock	Ptbn	925	43	/
YLZ-ZK-1	Carbonaceous shale	Nhd ₁	925	43	/
YLZ-ZK-1	Gray clay rock	Ptbn	650	45	/
Total			17,850	455	39.231

**Figure 12.** An example of K-mean clustering of the Mn/Cr ratio in a borehole in the Yanglizhang Deposit, Songtao ore-concentrated area.

In addition to the Mn/Cr ratio, audio magnetotelluric (AMT) detection can be useful in identifying hidden ore bodies in northeastern Guizhou. Traditionally, this method has been applied to explore the structures of ore-bearing level-III grabens and their associated level-IV grabens, and to predict whether a rhodochrosite layer is present in the level-IV grabens [50]. Although quantitative forward and inversion methods have been used to analyze the electrical structure characteristics of existing AMT detection profiles, qualitative interpretation methods are still used in the identification of structural frameworks and prediction of hidden ore bodies. Taking the prediction of prospective Mn ore deposits as an example, this paper uses logarithmic function fitting (Figure 13) to establish a logarithmic function prediction model for resistivity (ρ) and metallogenic intensity (φ):

$$\varphi = 1000(\ln(\rho) - 7.234177)/1351 \quad (8)$$

where metallogenic intensity (φ) equals the product of ore bed thickness (h) and grade (g).

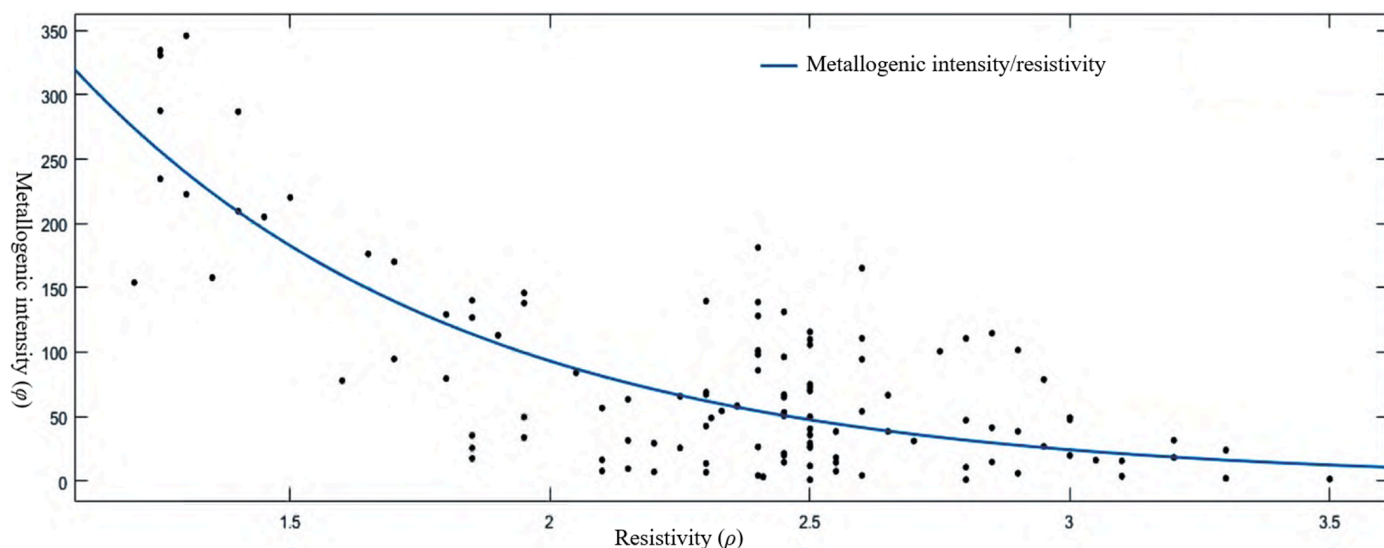


Figure 13. Spatial distribution and fitting curve of AMT-based resistivity (ρ) and metallogenetic intensity of rhodochrosite (φ = thickness $h \times$ grade g).

Equation (8) was tested ($F = 2.3974 \times 10^4$, $\alpha = 0.05$) and can indicate prospective ore body sections (5P) of Mn deposits.

From the multi-scale progressive prediction results for Mn ore deposits described above, it can be seen that geological data have big data attributes, and that geological work has data-intensive characteristics [42,49,62]. These form the basis of prospecting prediction using the fourth paradigm and the joint data-and-model-driven approach adopted in this study.

Big data not only comprise data for exploring intelligent prospecting modes but also represent an ideological and methodological foundation. Prospecting prediction based on geological big data includes analysis of metallogenetic backgrounds, ore-controlling factors, metallogenetic systems, metallogenetic sequences, and ore deposit distribution, and requires the adoption of a systematic, multilevel, and orderly approach. Our results demonstrate that because of the extreme complexity of the ore-forming process and its controlling factors, prospecting prediction based on big data must follow a logical process of mineral prospecting and exploration, combining model-driven and data-driven model-free methods, shifting from an area to a point and from the regional scale to the local scale. Then, considering the scale and accuracy requirements, the structured, unstructured, and semi-structured geological–geophysical–geochemical–remote-sensing data should be transformed into quantitative data, forming a complete data chain via assimilation and fusing. Finally, various targeted data mining algorithms should be organized into a digital, sequential, multidirectional, and multilevel technical methodology system to complete the gradual approximation from the periphery to the interior, gradually narrowing the scope, and progressively predicting prospective areas at a large (1P) to small scale (5P).

Based on a multi-scale progressive intelligent prospecting prediction method, combined with geological–geophysical–geochemical–remote-sensing data from the study area, we delineated a potential exploration target (the pink polygon in Figure 10). Drilling verification confirmed the presence of a hidden Mn ore deposit with a thickness of 5.47 m and an average grade of 18.66%, including a rich Mn ore deposit with a thickness of 1.67 m and an average grade of 27.25% [41,55]. The newly added manganese resources amount to 23.08 million tonnes, and the deposit reaches a large scale [41,55]. The methodology proposed in this paper provides a practical technical solution for the efficient exploration of deep super-large manganese deposits in northeastern Guizhou.

5. Conclusions

The main conclusions of the study are as follows:

(1) Optimizing the prospecting process for hidden DMO Mn ore deposits and integrating relevant prospecting knowledge allows for model-based prediction founded on basic ore mineralogy knowledge and model-free prediction based on the fourth paradigm. Consequently, big datasets and big data chains can be constructed, and a data-and-model-driven BDMPM method for multi-scale progressive prediction of hidden Mn ore deposits can be established using AI methods.

(2) In northeastern Guizhou, the Nanhua System in the Yanshanian Songtao–Rongjiang foreland depression belt is a probable Mn ore-forming area (1P). Areas where the ENE-trending Nanhua Rift secondary tectonic unit and the NW-trending ancient rift co-occur are permissive ore-finding areas (2P). Areas with geological, geophysical, geochemical, and remote-sensing comprehensive anomalies are (after elimination of interference from overburden and faults) preferable ore-finding areas (3P). Spectral anomalies of symbiotic and associated (secondary) Mn minerals can reveal potential mineral resource areas (4P). Clear and reliable physicochemical information reflecting the presence of ore bodies, such as the Mn/Cr ratio and the logarithmic relationship between metallogenic intensity and resistivity, can indicate prospective ore body areas (5P).

(3) The validation of BOA-AdaBoost in the Songtao Mn-ore-concentrated area demonstrated that the prediction model at the 3P scale achieved an accuracy of 0.936, precision of 0.948, recall of 0.925, F1 score of 0.928, Kappa of 0.862, and AUC of 0.967. Drilling verification confirmed the presence of a hidden Mn ore deposit with a thickness of 5.47 m and an average grade of 18.66%, including a rich Mn ore deposit with a thickness of 1.67 m and an average grade of 27.25% [41,55]. The newly added Mn resources amount to 23.08 million tonnes, and the deposit reaches a large scale [41,55]. This demonstrates that the proposed BDMPM provides a practical technical solution for the efficient exploration of deep super-large manganese deposits in northeastern Guizhou.

(4) The joint data-and-model-driven approach to prospecting prediction is complex because of the nature of big data: the data comprise geological, geophysical, geochemical, and remote-sensing datasets, encompassing multiple sources, types, dimensions, quantities, scales, time points, and themes [3,63]. Further research on geological big data mining and AI algorithms is required. Future work should not only focus on certain individual parameters and single algorithms, but also include computer application research on individual parts of the process. In addition, it should emphasize strengthening the overall logical process and data chain organization for data-driven prospecting prediction.

(5) The BDMPM has been validated on the DMO Mn ore deposits in northeastern Guizhou; in the future, it will be tested on lead–zinc deposits in northwestern Guizhou and barite deposits in southeastern Guizhou to verify its universality and robustness.

In conclusion, by reviewing existing successful cases, summarizing the knowledge of mineral exploration in the study area, drawing on the valuable experience of predecessors, and establishing a data-and-model-driven prospecting prediction model, this study demonstrated an effective method for mineral exploration.

Author Contributions: Conceptualization, K.X. and C.W.; methodology, K.X. and C.K.; software, C.K.; validation, C.K. and J.M.; formal analysis, C.K. and J.M.; investigation, K.X. and C.K.; resources, C.K.; data curation, C.K.; writing—original draft preparation, K.X.; writing—review and editing, C.K. and K.X.; visualization, J.M.; supervision, C.W.; project administration, C.W. and K.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will be made available on request.

Acknowledgments: We thank the Deep Earth Probe and Mineral Resources Exploration National Science and Technology Major Project (2025ZD1006500); Guizhou Provincial Scientific and Technological Program (QKBPT SYSS [2026]002, QKHPT SSYS[2026]ZD025, and QKHZD[2025]016); Open Research Project of the Hubei Key Laboratory of Intelligent Geo-Information Processing (KLIGIP-2023-B08); and the Detailed Survey of Key Mineral Resources in Guizhou Province: Detailed Survey Project of Manganese Ore Deposit in Songtao County. We are grateful to the reviewers for their valuable comments on this article.

Conflicts of Interest: The authors declare no competing interests.

References

1. Sun, T.; Feng, M.; Pu, W.; Liu, Y.; Chen, F.; Zhang, H.; Huang, J.; Mao, L.; Wang, Z. Fractal-based multi-criteria feature selection to enhance predictive capability of AI-driven mineral prospectivity mapping. *Fractal Fract.* **2024**, *8*, 224. [\[CrossRef\]](#)
2. Zhang, H.; Sun, T. Bibliometric analysis of mineral prospectivity mapping research from 2006 to 2024. *Nat. Resour. Res.* **2026**, *35*, 137–163. [\[CrossRef\]](#)
3. Zhou, Y.; Guo, Y. To readers from special editor in chief: A review of advances in big data and AI for geoscience. *Earth Sci. Front.* **2026**, *33*, 1–11.
4. Xu, Y.; Zuo, R. An interpretable graph attention network for mineral prospectivity mapping. *Math. Geosci.* **2024**, *56*, 169–190. [\[CrossRef\]](#)
5. Daviran, M.; Maghsoudi, A.; Ghezelbash, R. Optimized AI-MPM: Application of PSO for tuning the hyperparameters of SVM and RF algorithms. *Comput. Geosci.* **2025**, *195*, 105785. [\[CrossRef\]](#)
6. Boadi, B.; Sunder Raju, P.V.; Wemegah, D.D. Analysing multi-index overlay and fuzzy logic models for lode-gold prospectivity mapping in the Ahafo gold district-Southwestern Ghana. *Ore Geol. Rev.* **2022**, *148*, 105059. [\[CrossRef\]](#)
7. Xi, Y.; Li, Y.; Liu, J.; Wu, S.; Lu, N.; Liao, G.; Wang, Q. Application of analytic hierarchy process in mineral prospecting prediction based on an integrated geology-aerogeophysics-geochemistry model. *Minerals* **2023**, *13*, 978. [\[CrossRef\]](#)
8. Gajere, J.N.; Adekeye, O.A.; Tende, A.W.; Aminu, M.D. GIS-based multi-criteria decision models for barite exploration in Nigeria's Benue Trough. *Sci. Rep.* **2024**, *14*, 14031. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Yousefi, M.; Carranza, E.J.M.; Kreuzer, O.P.; Nykänen, V.; Hronsky, J.M.A.; Mihalasky, M.J. Data Analysis Methods for Prospectivity Modelling as Applied to Mineral Exploration Targeting: State-of-the-Art and Outlook. *J. Geochem. Explor.* **2021**, *229*, 106839. [\[CrossRef\]](#)
10. Yin, J.; Li, N. Ensemble learning models with a Bayesian optimization algorithm for mineral prospectivity mapping. *Ore Geol. Rev.* **2022**, *145*, 104916. [\[CrossRef\]](#)
11. Carranza, E.J.M.; Laborte, A.G. Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of random forests algorithm. *Ore Geol. Rev.* **2015**, *71*, 777–787. [\[CrossRef\]](#)
12. Zuo, R.; Cheng, Q.; Xu, Y.; Yang, F.; Xiong, Y.; Wang, Z.; Kreuzer, O.P. Explainable artificial intelligence models for mineral prospectivity mapping. *Sci. China Earth Sci.* **2024**, *67*, 2864–2875. [\[CrossRef\]](#)
13. Harris, J.R.; Strong, J.; Thurston, P.; Nymoen, K.; Haugaard, R.; Naghizadeh, M.; Tuba, G.; Behnia, P.; Grunsky, E.; Ayer, J. Mineral prospectivity mapping and differential metal endowment between two greenstone belts in the Canadian Superior Craton. *Nat. Resour. Res.* **2025**, *34*, 97–120. [\[CrossRef\]](#)
14. Agterberg, F.P. Computer programs for mineral exploration. *Science* **1989**, *245*, 76–81. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Agterberg, F.P. A modified weights-of-evidence method for regional mineral resource estimation. *Nat. Resour. Res.* **2011**, *20*, 95–101. [\[CrossRef\]](#)
16. Ngowi, D.E.; Mshiu, E.E.; Msabi, M. Mineral prospectivity modelling to delineate potential areas for gold deposits: A case study of Lupa Goldfield, South West Tanzania. *Appl. Earth Sci.* **2023**, *132*, 236–251. [\[CrossRef\]](#)
17. Zuo, R.; Wang, Z. Effects of random negative training samples on mineral prospectivity mapping. *Nat. Resour. Res.* **2020**, *29*, 3443–3455. [\[CrossRef\]](#)
18. Xiao, F.; Chen, W.; Wang, J.; Erten, O. A hybrid logistic regression: Gene expression programming model and its application to mineral prospectivity mapping. *Nat. Resour. Res.* **2021**, *31*, 2041–2064. [\[CrossRef\]](#)
19. Li, Q.; Chen, G.; Wang, D. Mineral prospectivity mapping using semi-supervised machine learning. *Math. Geosci.* **2025**, *57*, 275–305. [\[CrossRef\]](#)
20. Li, Y.; Fu, C.; Song, Y.; Zhang, Y.; Li, J.; Liu, Y.; Huang, B.; Li, Y. Integrating cost-sensitive ensemble learning and geoscience big data for Uranium prospectivity mapping in the Husab Uranium Deposit, Namibia. *Nat. Resour. Res.* **2026**, *35*, 3051–3070. [\[CrossRef\]](#)
21. Yu, Z.; Li, B.; Wang, X. Mineral prospectivity mapping susceptibility evaluation based on interpretable ensemble learning. *Ore Geol. Rev.* **2024**, *173*, 106248. [\[CrossRef\]](#)

22. Chen, M.; Xiao, F. Projection pursuit random forest for mineral prospectivity mapping. *Math. Geosci.* **2023**, *55*, 963–987. [\[CrossRef\]](#)
23. Qaderi, S.; Maghsoudi, A.; Yousefi, M.; Pour, A.B. Translation of mineral system components into time step-based ore-forming events and evidence maps for mineral exploration: Intelligent mineral prospectivity mapping through adaptation of recurrent neural networks and random forest algorithm. *Ore Geol. Rev.* **2025**, *179*, 106537. [\[CrossRef\]](#)
24. Bergen, K.J.; Johnson, P.A.; de Hoop, M.V.; Beroza, G.C. Machine learning for data-driven discovery in solid Earth geoscience. *Science* **2019**, *363*, eaau0323. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Li, C.; Xiao, K.; Sun, L.; Tang, R.; Dong, X.; Qiao, B.; Xu, D. CNN-Transformers for mineral prospectivity mapping in the Maodeng-Baiyinchagan area, Southern Great Xing'an Range. *Ore Geol. Rev.* **2024**, *167*, 106007. [\[CrossRef\]](#)
26. Hoseinzade, Z.; Shojaei, M.; Khademi, F.; Mokhtari, A.R.; Saremi, M. Integration of deep learning models for mineral prospectivity mapping: A novel Bayesian index approach to reducing uncertainty in exploration. *Model. Earth Syst. Environ.* **2025**, *11*, 161. [\[CrossRef\]](#)
27. Xiong, Y.; Zuo, R. A positive and unlabeled learning algorithm for mineral prospectivity mapping. *Comput. Geosci.* **2021**, *147*, 104667. [\[CrossRef\]](#)
28. Yin, B.; Zuo, R.; Xiong, Y. Mineral prospectivity mapping via gated recurrent unit model. *Nat. Resour. Res.* **2022**, *31*, 2065–2079. [\[CrossRef\]](#)
29. Chen, Q.; Cui, Z.; Liu, G.; Yang, Z.; Ma, X. Deep convolutional generative adversarial networks for modeling complex hydrological structures in Monte-Carlo simulation. *J. Hydrol.* **2022**, *610*, 127970. [\[CrossRef\]](#)
30. Lin, Q.; Zuo, R. Domain adversarial neural network for mapping mineral prospectivity. *Math. Geosci.* **2025**, *57*, 471–498. [\[CrossRef\]](#)
31. Lauzon, D.; Gloaguen, E. Quantifying uncertainty and improving prospectivity mapping in mineral belts using transfer learning and random forest: A case study of copper mineralization in the Superior Craton Province, Quebec, Canada. *Ore Geol. Rev.* **2024**, *166*, 105918. [\[CrossRef\]](#)
32. Shi, Z.; Zuo, R.; Zhou, B. Deep reinforcement learning for mineral prospectivity mapping. *Math. Geosci.* **2023**, *55*, 773–797. [\[CrossRef\]](#)
33. Xiong, Y.; Zuo, R.; Carranza, E.J.M. Mapping mineral prospectivity through big data analytics and a deep learning algorithm. *Ore Geol. Rev.* **2018**, *102*, 811–817. [\[CrossRef\]](#)
34. Zuo, R.; Xiong, Y.; Wang, Z.; Wang, J.; Kreuzer, O.P. A new generation of artificial intelligence algorithms for mineral prospectivity mapping. *Nat. Resour. Res.* **2023**, *32*, 1859–2186. [\[CrossRef\]](#)
35. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Deep learning and process understanding for data-driven earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#)
36. Zuo, R.; Yang, F.; Cheng, Q.; Kreuzer, O.P. A novel data-knowledge dual-driven model coupling artificial intelligence with a mineral systems approach for mineral prospectivity mapping. *Geology* **2025**, *53*, 284–288. [\[CrossRef\]](#)
37. Kong, Y.; Chen, G.; Liu, B.; Xie, M.; Yu, Z.; Li, C.; Wu, Y.; Gao, Y.; Zha, S.; Zhang, H.; et al. 3D mineral prospectivity mapping of Zaozigou gold deposit, west Qinling, China: Machine learning-based mineral prediction. *Minerals* **2022**, *12*, 1361. [\[CrossRef\]](#)
38. Garnit, H.; Kraemer, D.; Bouhlel, S.; Davoli, M.; Barca, D. Manganese ores in Tunisia: Genetic constraints from trace element geochemistry and mineralogy. *Ore Geol. Rev.* **2020**, *120*, 103451. [\[CrossRef\]](#)
39. Chen, Q.; Liu, G.; Ma, X.; Que, X. Spatial Analysis. In *Encyclopedia of Mathematical Geosciences*; Sagar, D., Cheng, B.S., McKinley, Q., Agterberg, J., Eds.; Springer Nature: Cham, Switzerland, 2023.
40. Wu, C.; Liu, G.; Zhang, X.; Xu, K. Fundamental problems of integrated application of big data in Geoscience. *Bull. Geol. Sci. Technol.* **2020**, *39*, 1–11. [\[CrossRef\]](#)
41. Kong, C.; Zhao, J.; Li, B.; Wu, C.; Xu, K. Manganese mineral prospectivity mapping based on semi-supervised learning and multi-source Geoscientific-sample Wasserstein Generative Adversarial Network (Geo-WGAN) in Songtao of Guizhou, South China. *Ore Geol. Rev.* **2025**, *186*, 106933. [\[CrossRef\]](#)
42. Wu, C.; Liu, G.; Zhang, X.; He, Z.; Zhang, Z. Discussion on Geological Science Big Data and its Applications. *China Sci. Bull.* **2016**, *61*, 1797–1807. [\[CrossRef\]](#)
43. Gray, J. Jim Gray on eScience: A transformed scientific method. In *The Fourth Paradigm*; Hey, A.J.G., Tansley, S., Tolle, K.M., Eds.; Microsoft Research: Cambridge, UK, 2009; Based on the Transcript of a Talk Given by Jim Gray to the NRC-CSTB in Mountain View, CA, 11 January 2007.
44. Mayer-Schönberger, V.; Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*; Houghton Mifflin Harcourt: New York, NY, USA, 2013.
45. Wu, C.; Zhang, S.; Xu, K.; Yuan, L.; Kong, C.; Zhang, Z.; Shen, H. Sedimentary evolution for the Wuling Secondary Rift of the western Nanhua Rift and exhalative-sedimentary mineralization cycles of “Datangpo-type” manganese ore deposits in South China. *J. Asian Earth Sci.* **2025**, *287*, 106566. [\[CrossRef\]](#)
46. Zhang, S.; Yuan, L.; Long, J.; Liu, Y.; Pan, W.; Shen, H. Genesis and role of Bubble-like structures in Rhodochrosite-rich orebodies of “Datangpo-type” sedimentary exhalative manganese deposits in China. *J. Asian Earth Sci.* **2025**, *292*, 106741. [\[CrossRef\]](#)

47. Zhou, Q.; Wu, C.; Hu, X.; Yang, B.; Zhang, X.; Du, Y.; Xu, K.; Yuan, L.; Ni, J.; Hu, D.; et al. A New Metallogenic Model for the Giant Manganese Deposits in Northeastern Guizhou, China. *Ore Geol. Rev.* **2022**, *149*, 105070. [\[CrossRef\]](#)
48. Zhao, P.; Chi, S.; Chen, Y. A thorough investigation of geo-anomaly: A basis of metallogenic prognosis. *Geo. J. China Univer.* **1996**, *4*, 361–373.
49. Zhao, P. “Three-component” quantitative resource prediction and assessments: Theory and practice of digital mineral prospecting. *Earth Sci.* **2002**, *27*, 482–489.
50. Xu, K.; Wu, C.; Zhang, S.; Ma, X.; Yang, B.; Kong, C. Extraction of NW-trending, ore-conducting basement hidden faults in manganese ore concentration area based on multi-source data in Northeastern Guizhou, China. *Minerals* **2026**, *16*, 58. [\[CrossRef\]](#)
51. Du, Y.; Yu, W.; Zhou, Q.; Liu, Z.; Zhang, Z.; Wang, P.; Guo, H.; Xu, L.; Lv, S.; Huang, H.; et al. Hydrothermal exhalative manganese deposits in Chinese rift basins and their metallogenic mechanism[OL]. *Geotectonica et Metallogenia* **2026**. Available online: <https://link.cnki.net/urlid/44.1595.p.20260904.1542.002> (accessed on 7 September 2026).
52. Yang, B.; Hu, X.; Gao, L.; Shen, X.; Liao, W.; Wang, J.; Yuan, L. Discharge and deposition metallogenic model of “Datangpo-type” manganese deposits in China: Geophysical evidence from deep electrical structure of manganese-bearing grabens. *J. Asian Earth Sci.* **2025**, *292*, 106732. [\[CrossRef\]](#)
53. Zhou, Q.; Yuan, L.; Yang, B.; Zhang, S.; Xu, K.; Xie, X.; Zhang, X.; Pan, W. Development and application of a technological system for exploration of “Datangpo-type” concealed manganese deposits in China. *J. Asian Earth Sci.* **2025**, *292*, 106727. [\[CrossRef\]](#)
54. Zhou, Q.; Liu, Y. Metallogenic model, exploration methods, and prospecting progress of exhalative sedimentary manganese deposits. *J. Chengdu Univ. Technol.* **2026**, *53*, 725–740.
55. Zhou, Q.; Wu, C. Experimental Research on Big Data-based Intelligent Exploration Models and Advance. *Earth Sci. Front.* **2024**, *31*, 350–367.
56. Xu, K.; Zhao, S.; Wu, C.; Zhang, S.; Yuan, L.; Yang, C.; Li, Y.; Dong, Y.; Wu, Y.; Xiang, S.; et al. Manganese mineral prospectivity based on deep convolutional neural networks in Songtao of Northeastern Guizhou. *Earth Sci. Inform.* **2024**, *17*, 1681–1697. [\[CrossRef\]](#)
57. Chen, Y.L.; Sui, Y.H. Dictionary learning for integration of evidential layers for mineral prospectivity modeling. *Ore Geol. Rev.* **2022**, *141*, 104649. [\[CrossRef\]](#)
58. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* **2015**, *104*, 148–175. [\[CrossRef\]](#)
59. Garrido-Merchán, E.C.; Hernández-Lobato, D. Dealing with categorical and integer-valued variables in Bayesian optimization with Gaussian processes. *Neurocomputing* **2020**, *380*, 20–35. [\[CrossRef\]](#)
60. Hajihosseini, M.; Maghsoudi, A.; Ghezelbash, R. Development of a hybrid Bayesian-hyperband optimization procedure: GeoAI-driven hyperparameter tuning of AdaBoost for enhancing mineral prospectivity mapping. *Environ. Model. Softw.* **2026**, 106883. [\[CrossRef\]](#)
61. Xu, K.; Yuan, L.; Yang, B.; Kong, C.; Zhang, X.; Zheng, J.; Zhou, Q.; Wu, C. Extraction of hidden manganese ore information with combined mining of associated and secondary mineral of remote sensing data on Northeast Guizhou. *Bull. Geol. Sci. Technol.* **2020**, *39*, 37–43.
62. Zhao, P. Digital mineral exploration and quantitative evaluation in the big data age. *Geol. Bull. China* **2015**, *34*, 1255–1259.
63. Zhou, Y.; Xiao, F. Overview: A glimpse of the latest advances in artificial intelligence and big data geoscience research. *Earth Sci. Front.* **2024**, *31*, 1–6. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.